Synthetic and Systems Biotechnology 9 (2024) 540–548

Contents lists available at ScienceDirect

Synthetic and Systems Biotechnology

journal homepage: www.keaipublishing.com/en/journals/synthetic-and-systems-biotechnology

Original Research Article

Determination of the chromosomal position effects for plug-and-play application in the *Myxococcus xanthus* chassis cells

Xin-jing Yue **, Jia-rui Wang, Jun-ning Zhao, Zhuo Pan, Yue-zhong Li *

State Key Laboratory of Microbial Technology, Institute of Microbial Technology, Shandong University, Qingdao, PR China

ARTICLE INFO

Keywords:
Position effect
High expression integration sites (HEISs)
Myxococcus xanthus
Tn-seq

ABSTRACT

The chromosomal position effect can significantly affect the transgene expression, which may provide an efficient strategy for the inauguration of alien genes in new hosts, but has been less explored rationally. The bacterium *Myxococcus xanthus* harbors a large circular high-GC genome, and the position effect in this chassis may result in a thousand-fold expression variation of alien natural products. In this study, we conducted transposon insertion at TA sites on the *M. xanthus* genome, and used enrichment and dilution indexes to respectively appraise high and low expression potentials of alien genes at insertion sites. The enrichment sites are characteristically distributed along the genome, and the dilution sites are overlapped well with the horizontal transfer genes. We experimentally demonstrated the enrichment sites as high expression integration sites (HEISs), and the dilution sites unsuitable for gene integration expression. This work highlights that HEISs are the plug-and-play sites for efficient expression of integrated genes.

1. Introduction

"Position effect", reported in either eukaryotes or prokaryotes, refers to the phenomenon that the expression of a gene is altered by its location on chromosome [1]. A classic example of position effect is the translocation of the white gene of *Drosophila melanogaster* to heterochromatin, which causes the original solid red eye a white and red mottled appearance [2]. In mammals, the expression alteration of crucial genes derived from chromosome location changes is often highly relevant to the genetics of cancer and rare diseases [3]. The position effect has been experimentally confirmed by inserting reporter genes into the chromosome of mice [4], yeast [5,6], *Escherichia coli* [7,8] and *Bacillus subtilis* [9]. Position effect is also very important for the heterologous expression of large biosynthetic gene clusters of secondary metabolites (smBGCs), which is often a conundrum to solve. For example, the production of aranciamycin varied up to eightfold depending on the insertion sites of the biosynthetic genes in the genome of *Streptomyces albus* [10]. Similarly, random insertion of the epothilone gene cluster in the *Myxococcus xanthus* genome led to 0.1 μg/L to 1.2 mg/L of the epothilone production [11], and this production variation is greatly higher than subsequent multiple genetic improvements [12–15]. The position effect thus provides a potent means to achieve efficient expression of target genes or gene clusters in microbial cell factories, especially for their inauguration in new chassis, which, however, requires to screen lots of random insertion mutants. Up to date, the position effect has not been systematically and rationally explored for transgene expressions in any organisms withal.

M. xanthus, the model species of myxobacteria, is an ideal and the only myxobacterial chassis for the heterologous expression of myxobacteria-sourced genes like smBGCs, because of its phylogenetically close relationship, relatively fast growth ability and genetic maneuverability [16,17]. We previously inserted the chloramphenicol acetyl transferase (CAT) gene into the *M. xanthus* genome via site-specific recombination or random transposition, and found a stable expression with the site-directed integration but varied expressions with random insertions, especially for the large biosynthetic gene cluster of epothilones [11]. This is the first report of the position effect in myxobacteria, and the thousand-fold expression variation caused by position effect suggests its promising application for efficient expression of target genes in *M. xanthus* chassis. In this study, to characterize the position

Peer review under responsibility of KeAi Communications Co., Ltd.
* Corresponding author.
** Corresponding author.
E-mail addresses: xjy2018@sdu.edu.cn (X.-j. Yue), 202000141099@mail.sdu.edu.cn (J.-r. Wang), m15636092040@163.com (J.-n. Zhao), panzhuoddv@163.com (Z. Pan), lilab@sdu.edu.cn (Y.-z. Li).

https://doi.org/10.1016/j.synbio.2024.04.007
Received 27 November 2023; Received in revised form 30 March 2024; Accepted 8 April 2024
Available online 10 April 2024

effects in *M. xanthus*, we identified the global expression patterns of integrated alien genes with the transposon insertion sequencing (Tn-seq), and used enrichment and dilution indexes to respectively appraise high and low expression potentials of genes at insertion sites. The enrichment sites are characteristically distributed along the genome, and the dilution sites are overlapped well with the horizontal transfer genes. We experimentally demonstrated that these enrichment sites are high expression integration sites (HEISs), which are potential plug-and-play sites for efficient expression of integrated genes.

2. Materials and methods

2.1. Bacterial strains and culture conditions

M. xanthus DK1622 strain and mutants were grown in CTT medium [18] at 30 °C. *E. coli* was cultured in LB broth at 37 °C. The pBJ113 and deuterogenic plasmids were constructed in *E. coli* DH5α, and the pSWU19, pMiniHimar-lacZ, and deuterogenic plasmids were constructed in *E. coli* DH5α λ pir. When required, the LB and CTT media were supplemented with kanamycin (km; 40 μg/ml, 200 μg/ml or 1000 μg/ml). All the strains and plasmids used in this study are listed in Tables S1 and S2, respectively.

2.2. Transcriptome analysis

The *M. xanthus* DK1622 cells were collected after 28 h of incubation, followed with total RNA extraction with the TRIzol® Reagents according to the manufacturer's instructions (Invitrogen) and genomic DNA was removed using DNase I (Takara). The RNA quality was determined by 2100 Bioanalyser (Agilent) and quantified using the ND-2000 (NanoDrop Technologies). The RNA was used to prepare the RNA-seq transcriptome library with TruSeqTM RNA sample preparation Kit, which was sequenced on the Illumina HiSeq × TEN platform. Experimental details were described in our previous work [11]. Raw data was filtered with htqc [19], and the cleaned reads were mapped onto the *M. xanthus* DK1622 genome with SOAP [20].

2.3. Prediction and verification of the oriC of M. xanthus DK1622

The *oriC* region of *M. xanthus* DK1622 genome was predicted with Ori-Finder [21]. For verification, we replaced the replicon of pZJY41, a shuttle plasmid containing the replicon of the self-replicating plasmid pMF1 [22], with the DKoriC fragment and introduced the recombinant plasmid pZJY41-DKoriC into *M. xanthus* DK1622 to construct mutants, from which the pZJY41-DKoriC′ plasmids were extracted again to verify whether the DKoriC worked as an *oriC* to initiate the plasmid replication. In detail, the predicted *oriC* sequence was amplified from DK1622 genome using primers DK-oriC F/R, and then reconstructed into the linear pZJY41 lacking the pMF1 *oriC*, which was amplified with 41 F/R primers, to construct the pZJY41-DKoriC. The pZJY41-DKoriC plasmid was introduced into *M. xanthus* DK1622 by electroporation and the mutants DK-DKoriC were screened with kanamycin after 7 days of incubation. Electroporation and mutant screening was performed as previous report [13]. Total DNA (genome and plasmid) was extracted from mutants with Genome DNA extraction kit (Omega, USA) as the manufacturer's instructions, and the extracted DNAs were detected by gel electrophoresis. Because of the low copy number of plasmids in *M. xanthus*, leading to band invisibility of the extracted pZJY41-DKoriC, the extracted DNAs were further introduced into *E. coli* DH5α by heat shock transformation, and the clones resistant to kanamycin were picked and incubated to extract the pZJY41-DKoriC′ with Mini plasmid extraction kit (Tiangen, China). The extracted plasmids pZJY41-DKoriC′ were verified by gel electrophoresis, PCR with DK-oriC F/R and subsequent sequencing. The primers used were listed in Table S3.

2.4. Construction of transposable plasmids and M. xanthus mutant libraries

In Himar1 mariner transposon, the transposase recognizes the inverted repeats (IRs) sequence and integrates the sequence between the two IRs into the TA dinucleotide sites in the host genome [23]. Based on the Himar1 mariner transposon in plasmid pminiHimar-lacZ, a single nucleotide replacement in both (IR) changes the IR sequence from 5′-agaccggggacttatcaGccaacctgt-3′ to 5′-agaccggggacttatcaTccaacctgt -3' (the *Mme*I site is underlined and the mutation is in uppercase), thereby introducing the *Mme*I restriction site in pminiHimar-Y. Using pminiHimar-Y as template, the fragment Tpase-IR1, km-IR2, and R6K were amplified with corresponding primer pairs, and then regrouped into a simplified transposition plasmid Ymini with ClonExpress II One Step Cloning Kit (Vazyme, China). Two transcriptional terminators (TTs) were amplified with rrnB-T F/R primer pairs from plasmid pkk232-8 [24], and then cloned into Ymini, resulting Ymini2. Subsequently, the *egfp* controlled by the BBa_J23100 promoter was amplified from pBJ115-*egfp*-J23100-ddva with the BJ-*egfp* F/R primer pair and introduced into the region between two TTs in Ymini2 to construct Ymini3. All constructs were verified by sequencing.

In the transposon plasmid Ymini3, the *egfp* reporter gene was controlled by a strong promoter BBa_J23100, while the *km*^{r} was under the control of a moderate Tn5 promoter [25]. Two TTs were constructed upstream and downstream the *egfp* gene to isolate the transcription of two reporter genes, as well as the read-through from the in-situ genes. Ymini3 was introduced into *M. xanthus* DK1622 by electroporation as previously reported [13] with minor modifications. The *M. xanthus* cells were harvested after 24 h of incubation, washed with cold ddH_2O three times, and then resuspended to 50 OD. 100 μL of the cell suspension was mixed with 100 μg of Ymini3 and electroporated at a voltage of 1250 V in a 2-mm cuvette using the Electroporator (Eppendorf, Germany). After being introduced into *M. xanthus* DK1622, the IR-*egfp*-*km*^{r}-IR (GK cassette) was cut from Ymini3 and integrated at the TA site of DK1622 genome [26], resulting the transposable mutant. Screening based on the *egfp* expression was preferred, because the mutants had to be mixed with soft agar on screening plates due to the growth characteristics of myxobacteria, which made it almost impossible to process the library samples into homogeneous suspension for the screening by flow cytometry. Therefore, the mutants were screened on plates containing 40, 200, and 1000 μg/ml of kanamycin, respectively, to construct the libraries km40, km200, and km1000. Mutant clones were collected from multiple rounds of electrical transformation for subsequent Tn-seq analysis. The primers used were listed in Table S3.

2.5. Tn-seq analysis

The mutant clones of different libraries were scraped and mixed from screening plates for the extraction of mix genome DNAs. The mutant cells were resuspended with 5 ml of TE buffer (25 mM Tris-HCl, 25 mM EDTA, pH 8.0) containing 0.5% SDS and 100 μl of lysozyme. After an incubation at 37 °C for 1 h, 25 μl of protease K were added and incubated for another 1 h. An equal volume of tris-phenol was added to denature proteins, which were then removed by centrifugation. The supernatant was mixed with 10% of 3 M NaAC and an equal volume of cold isopropyl alcohol, then stored at −20 °C for 30 min. The mixed genome DNA was collected after centrifugation at 3000 rpm for 20 min and washed with 6 ml of cold 70% ethanol. The DNA was vacuum-dried and dissolved in 200 ml of TE buffer.

Preparation and sequencing of transposon population libraries were performed as the description of Opijnen TV and Camilli A [27] with minor modifications. The mixed genome DNA was digested with *Mme*I at 37 °C for 4 h, followed by dephosphorylation with CIP at 37 °C for 1 h, and then the extraction with phenol:chloroform:isoamyl alcohol, the precipitation with 95% ethanol, washing with 70% ethanol, and finally the resuspension in distilled water. The AP-A and AP-B primers were

shocked at 96 °C for 2 min and then cool to room temperature (~10–20 min) to prepare the adapters, which were ligated with the extracted genome DNA with T4 DNA ligase overnight at 16 °C. With the linked products as templates, the 120-bp target fragments were captured and amplified with primers P1_M6_MmeI and Gex PCR Primer 2, purified with DNA Purification Kit (Tiangen, China) and then sequenced using Illumina Nova 6000 platform. The target reads were screened according to the known sequence of IR and adapter, and aligned with the genome sequence with a 100% matching threshold. The primers used were listed in Table S3.

To compare the insertion frequency of each site and the relative abundance of the corresponding mutant in libraries, the insertion index of each site was calculated as read number of one site/total read number × 100000. Similarly, the insertion index of a gene was calculated as read number of a gene/number of theoretical TA sites/total read number × 100000. The insertion sites or genes having insertion indexes increasing by more than 2 times with increasing kanamycin concentrations for the screening were defined as enrichment sites or enriched genes. When screening the enrichment sites or enriched genes, those with less than 5 reads were removed [23]. According to the insertion indexes in km40, the top 1% sites (390 sites) with relatively high abundance were singled out, of which 262 sites were diluted by more than 2 times or even not inserted in km200 and/or km1000 libraries and therefore defined as dilution sites. The information of enrichment sites and dilution sites is listed in Table S4.

2.6. Construction of site-integrated plasmid and M. xanthus mutants

The *egfp* gene controlled by Tn5 promoter was amplified from pSWU19-*egfp* with primer pair *egfp*-F/R and then cloned between the two transcription terminators in plasmid pBJ-T, resulting pBJ-T-*egfp*. The 800-bp regions upstream of each insertion sites were amplified with the corresponding primer pairs as homologous arms and introduced into pBJ-T-*egfp* to construct a series of vectors pBJ-T-*egfp*-SN (N means the selected insertion site), which were introduced into *M. xanthus* DK1622 by electroporation, as described previously [14]. After screening with kanamycin, the mutants in which the *egfp* were integrated upstream of insertion sites by homologous recombination was selected and verified by colony PCR with primer pair *egfp*-F/R. Similarly, the pSWU19-*egfp* plasmid was introduced into *M. xanthus* DK1622, followed with integration of *egfp* into the attB site through site-specific recombination and construction of mutant DK-19-*egfp*.

The *idgS* gene were amplified from pCIMt002 [28] with primer pair idgs-F/R, and then ligated with the linear pBJ-T-SN amplified with primer pair pBJ117-F/R from plasmids pBJ-T-*egfp*-S1- pBJ-T-*egfp*-S5, resulting plasmids pBJ-T-*idgS*-S1- pBJ-T-*idgS*-S5. To replace the *egfp* in pSWU19-*egfp* and construct the pSWU19-*idgS*, the Tn5-*idgs*-F/R primers were used to amplify Tn5-*idgs* fragment from pBJ-T-*idgS*-S1, which was then ligated with the linear pSWU19 amplified with 19-*idgs* F/R primers from pSWU19-*egfp* plasmid. All plasmids were verified by sequencing, and then introduced into *M. xanthus* DK1622 to construct mutant DK-*idgS*-1- DK-*idgS*-5, and DK-19-*idgS*. The primers used are listed in Table S3.

2.7. Detection of the expression of egfp and idgS reporter genes

With primers *gfp* QF/QR, the transcription of *egfp* was analyzed with RT-qPCR as described previously [14]. The OD_{600} and fluorescence (excitation at 485 nm and emission at 526 nm) of samples were analyzed by UNICO-7200™ Spectrophotometer (USA) and a F-4600™ Fluorescence Spectrophotometer (Japan), respectively, and the fluorescence/OD_{600} values were used for measurement of the expression levels of *egfp*. Detailed methods refer to the previous report [13].

M. xanthus DK1622 and indigoidine-producing mutants were cultured in 50 mL of CTT medium (40 μg/mL of kanamycin was supplemented for mutants) at 200 rpm and 30 °C overnight, and then transformed into 50 mL of CTT medium (with three biological parallels) with a 2% inoculum size. After 48 h of incubation, 1 mL of the bacterial suspension was harvested by centrifugation at 8000×*g* for 5 min. 200 μL of the supernatant was transferred to a 96-well plate, and the absorbance was scored at 612 nm using UNICO-7200™ Spectrophotometer (USA), relative to a blank control containing 200 μL of CTT medium. The radio of OD_{612} to OD_{600} was used to measure the production of indigoidine.

2.8. Annotation and visualization of the DK1622 genome

The circular genome of *M. xanthus* DK1622 was annotated and visualized with the online proksee (https://proksee.ca/). The horizontal transfer genes were identified with the built-in Alien_hunter tool. The cumulative GC profile of DK1622 was analyzed with GC-Profile (https://tubic.org/CpG-Profile/public) [29].

3. Results

3.1. Gene organization and expression in M. xanthus genome

The *M. xanthus* DK1622 harbors a 9.14-Mb circular chromosome (NC_008095) with the 68.9% GC content, possessing 7232 protein-encoding genes (Fig. 1a and b). The chromosomal origin *oriC* (Fig. 1a) was predicted with the Ori-Finder [21]. To confirm the replication activity, we replaced the pMF1 replicon of the shuttle plasmid pZJY41 [22] with the DK1622 origin *oriC* fragment, which successfully imparts the hereditary stability to the recombinant plasmid pZJY41-DKoriC in *M. xanthus* (Fig. S1). The replication termination region (*Ter*) was predicted to locate at the region of 4.4–4.6 Mb (Fig. 1a), according to the *dif* sites [30].

In *Streptomyces*, the linear genome is separated by the first and last rRNA operons into a central core region comprising essential genes and two terminal chromosomal arms with genes coding for dispensable functions, which was termed “genetic partition” [31,32]. A similar genetic compartmentalization was found in the circular *M. xanthus* chromosome. There are four rRNA operons in *M. xanthus* DK1622 (the yellow bars under the horizontal histogram in Fig. 1a): two are near the *oriC*, while the other two, locating at 3.29 Mb and 5.89 Mb, divide the entire genome roughly into three equal parts. In the following random insertion experiment (see below), 669 genes, including the rRNA genes, were unable to be inserted and were defined as the essential genes (Table S5), which, however, scattered throughout the chromosome (Fig. 1c). Conversely, most of the 24 smBGCs in DK1622 [17] (Table S6) were clustered in the region (between the second and third rRNA operon) containing the *Ter* (Fig. 1d). The defined essential genes and smBGCs exhibited commensurately high and low transcriptional level, respectively (Fig. S2).

The transcriptome analysis of *M. xanthus* cells grown in CTT for 24 h revealed a common uneven peak chart of the endogenous genes (Fig. 1e). The median RPKM (reads per kilobase per million mapped reads) of exponential DK1622 cells was 24.6, and the top 10% high expression genes (HEGs, RPKM>215.7) mainly code for proteins involved in basic materials and energy metabolisms, while the 10% lowest expression genes (LEGs, RPKM<1.6) were significantly enriched in biotin metabolism, fatty acid synthesis, and secondary metabolic pathways (Figs. S3a and b). However, the HEGs and LEGs were apparently dispersed throughout the genome without the regional aggregation of functions.

3.2. Global position effects in the chromosome of M. xanthus

To explore the position effect on the chromosome scale in *M. xanthus* DK1622, we randomly integrated the GK reporter cassette containing the *egfp* and kanamycin resistance (km^r) genes into the dinucleotide TA sites of the genome to construct the random-insertion mutation libraries (Fig. 2). Different concentrations of kanamycin (40 μg/mL, 200 μg/mL

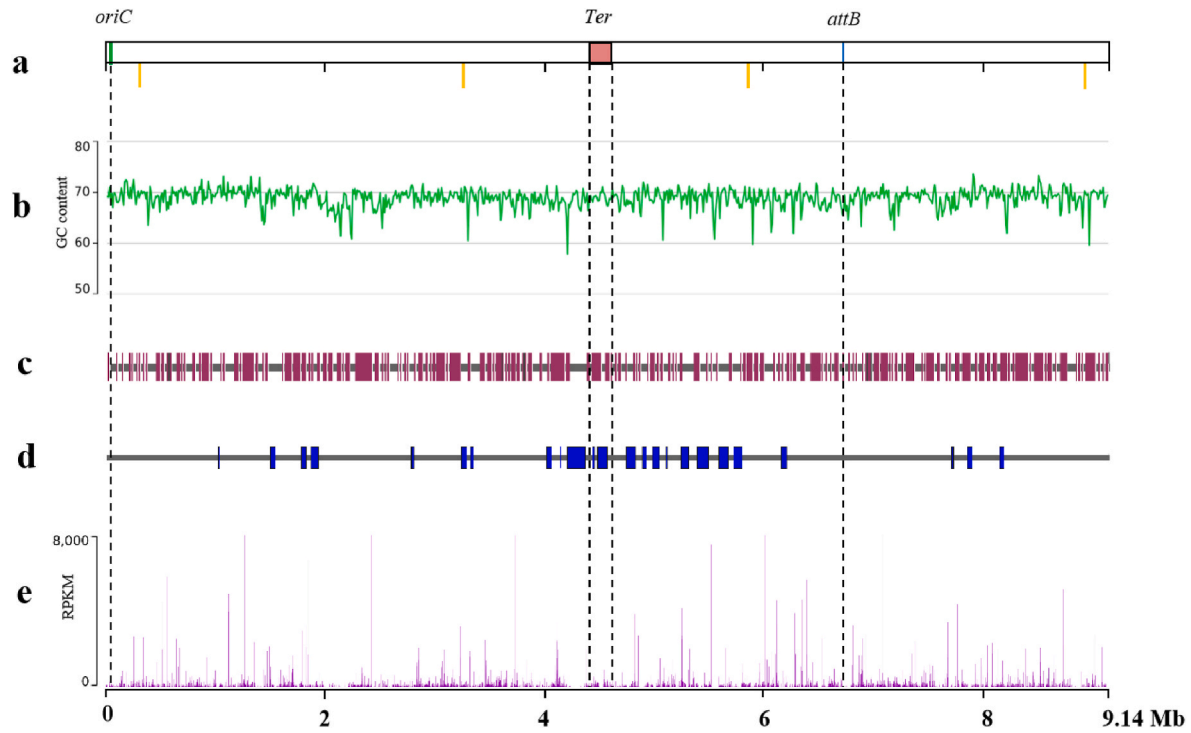

Fig. 1. The dimension and transcription of *M. xanthus* DK1622 chromosome. (a) The genetic compartmentalization of DK1622 chromosome. Green rectangle, *oriC* (21,268–21,643 bp); yellow rectangles, rRNA operons; red rectangle, replication terminal Ter; blue rectangle, *attB* site for site-specific recombination. (b) The GC content of *M. xanthus* DK1622 chromosome. (c) The distribution of essential genes in DK1622 chromosome. Red rectangles, genes not inserted in mutant libraries. (d) The distribution of smBGCs in DK1622 chromosome. (e) Global transcription of DK1622 chromosome. RPKM, reads per kilobase per million mapped reads.

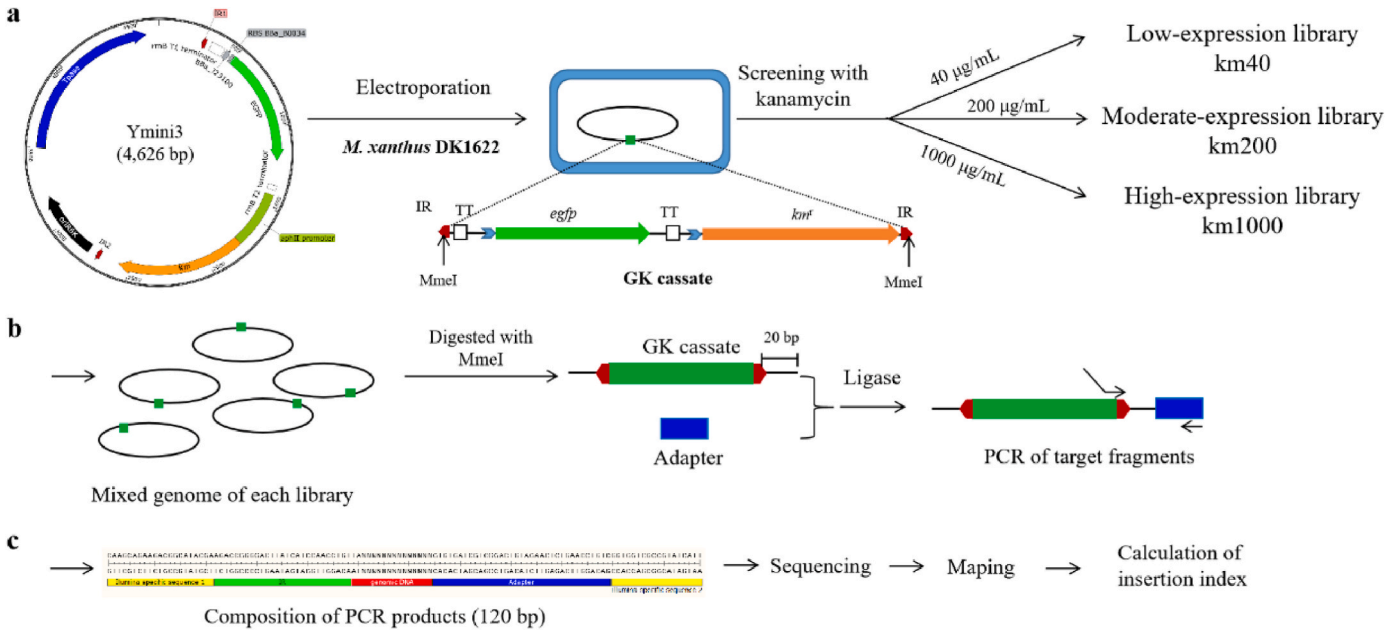

Fig. 2. Flowchart depicting Tn-seq, from library construction to massively parallel sequencing of transposon-chromosome junctions. (a) Construction of transposition mutation library of *M. xanthus* DK1622. The GK cassette integrated into DK1622 genome was marked with green square. IR, inverted repeat; TT, transcription terminator, km^r, kanamycin resistance gene. The mutants were screened into three libraries with different concentrations of kanamycin. (b) Preparation of library for Tn-seq sequencing. (c) Processing of Tn-seq sequencing data. Detailed information was described in Methods.

and 1000 μg/mL) were used to screen the mutants tolerating low, medium and high kanamycin selection pressures, which produced three mutant libraries, named as km40, km200 and km1000. Estimated from colony forming unit (CFU), 1 OD of the electroporated DK1622 cells (approximately 5×10^8 cells) produced ~12, 000 resistant mutant clones on the km40 screening plate. When the kanamycin concentration increased to 200 μg/mL and 1000 μg/mL, the number of colonies grown on the screening plates decreased by approximately 33% and 99%, respectively, which indicated that most of low-tolerance mutant strains were killed with the increase of kanamycin concentration, even though

the insertion might result in the expression obstruction of in situ genes.

After multiple rounds of electroporation, the mutant numbers were accumulated to more than 670,000, 240,000 and 15,000 in the km40, km200 and km1000 libraries, respectively. Theoretically, there are 282,395 TA sites allowing for the insertion of GK cassette. Because of the kanamycin selection pressure, we suggested that the total number of mutants used in the construction of each of the three libraries was all more than three times of the theoretic TA sites. Using the high-throughput sequencing, we obtained 1,933,139, 1,414,962 and 452,865 reads for the km40, km200 and km1000 libraries, respectively. The number of insertion sites was 56,324 in the km40, 56,046 in the km200 and 11,393 in the km1000 (Fig. 3a), which accounted for 19.9%, 19.8% and 3.7% of the total TA sites, respectively. We noticed that, even though the mutant number in the km40 library was 2.8 times of that in km200 library, the insertion coverage was not increased significantly in km40; whereas the mutant number in km1000 library was only approximately 6% of that in km200 library, but the insertion coverage increased three times, i.e., 18.7% of that in the km200 library. The results suggested that the high-tolerance insertion sites screened by the km1000 library almost reached saturation.

The distribution of inserted sites, as well as the relative abundance of mutant strains in three libraries, is shown as the histogram in Fig. 3b. Obviously, the insertion sites scattered across the *M. xanthus* DK1622 genome with no observed dense region in all the three libraries. Although the distribution of the TA sites was significantly lower in the gene coding regions than that in the intergenic regions (Fig. S4a, Wilcoxon test, $p < 0.001$), the same percentages (~4%) of the TA loci in these two regions were inserted in the km1000 library (Fig. S4b), suggesting no preference. We used insertion index to reflect the site insertion frequency to characterize the relative abundance of a site insertion in library. As shown in Fig. 3c, the median insertion index was improved from 1.32 in km40 to 1.48 in km200 and then 3.69 in km1000, suggesting that the loci for high resistance in km1000 was greatly focused. Examining the *M. xanthus* DK1622 encoding genes indicated that each of them contains at least a dinucleotide TA site. Totally, 6716 genes (93%) were inserted in the three libraries (Fig. S4c). The genes that could not be inserted were considered as essential genes (Table S5). We checked the top 10% of genes with high insertion indexes in the three libraries, and found no significant functional enrichment (Fig. S5), which suggested that the mutants with the high insertion abundance were not focused by functional inactivation of the in-situ genes.

3.3. Identification of HEISs in the M. xanthus chromosome

There were totally 6487 loci shared by all the km40, km200 and km1000 libraries (Fig. 3a). After removing the loci that had less than 5 reads, we screened the remaining 2945 loci to discern the potentials of having positive or negative position effects to the kanamycin tolerance of mutant strains. Among the 2945 loci, 1095 loci having 2-fold or higher increase of the insertion index with the screening pressure increase from 40 μg/mL to 200 μg/mL or from 200 μg/mL to 1000 μg/mL were defined as enrichment sites (details refer to Table S4), while 262 loci having a higher insertion index in km40 but lower or even zero insertion index in km200 and km1000 were considered as the dilution sites (Table S4, Fig. 4a).

Based on sequence characteristics of the insertion sites, the genomic positions of enrichment sites were classified into three types: within open reading frame (ORF), in regulatory intergenic region (RIR) and in nonregulatory intergenic region (NIR) (Fig. S6a). Totally, 719 local genes were inactivated by the insertion of GK cassette at enrichment sites, of which 60% had predictive functions, involving in various cellular metabolic processes with no significant enrichment (Figs. S6b and c), which again suggested that the enriched mutations with high tolerance against kanamycin were not produced from the influence of local genes or their regulatory factors, but high expression of km^r itself. In addition, transcription terminators were designed before the promoters of reporter genes in GK cassette (Fig. 2a), which also prevent the potential correlation between the insertion indexes of enrichment sites and the transcription levels of in situ genes (Fig. S6d). Therefore, the enrichment sites were suggested HEISs for efficient expression of integrated alien genes.

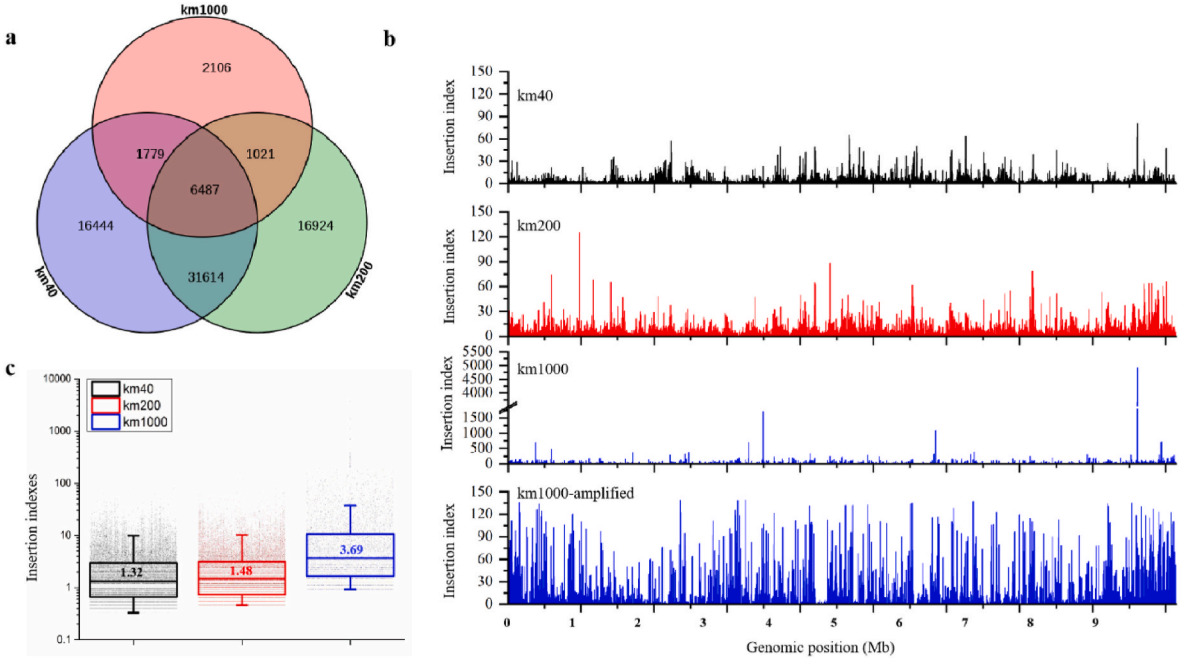

Fig. 3. Description of insertion sites in mutant libraries. (a) Venn diagram of the insertion sites in libraries km40, km200, and km1000. (b) Distribution of insertion sites with corresponding insertion index across the DK1622 genome in different libraries. The km1000 was also enlarged and displayed at the same scale of km40 and km200. (c) Comparation of the insertion index of each insertion site in different libraries. The median of insertion indexes was indicated.

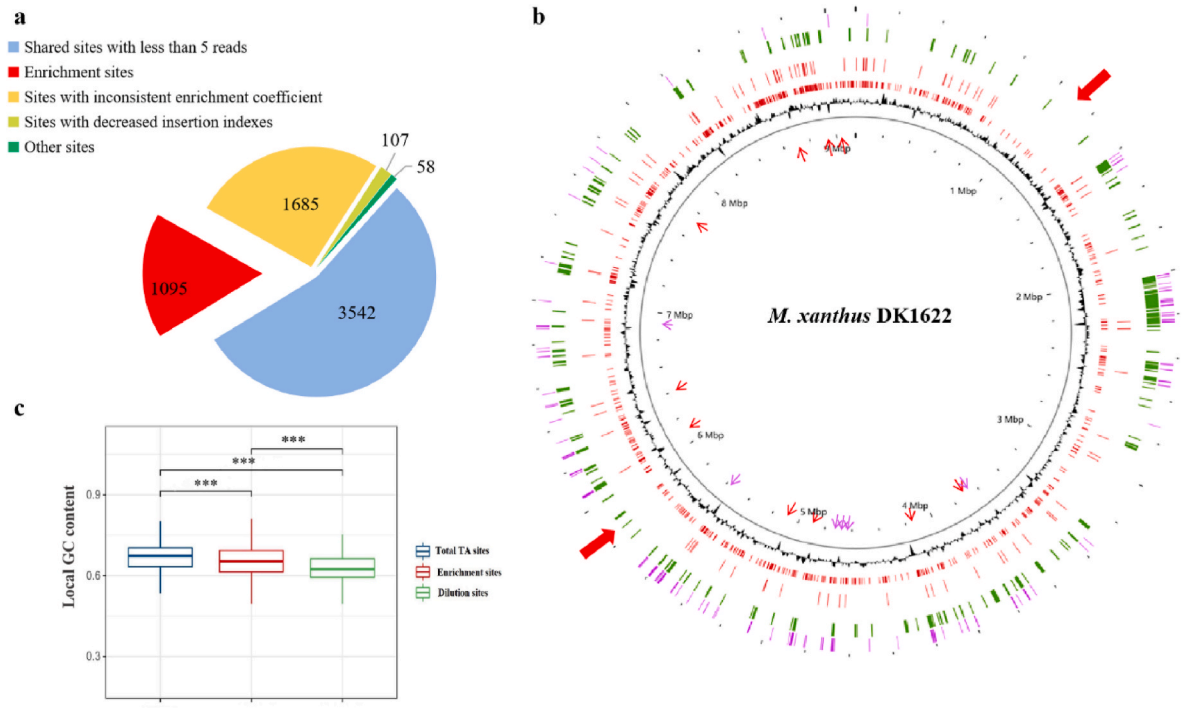

Fig. 4. Screening of enrichment sites. (a) Classification of insertion sites shared by three libraries. (b) Distribution of enrichment and dilution sites on the circular DK1622 genome. From the inside out, circle 1, genomic location information; circle 2 (black), GC content of DK1622 genome; circle 3 (red), all enrichment sites in DK1622 genome; circle 4 (red), the top 100 enrichment sites with large insertion index; circle 5 (green), horizontal transfer genes in DK1622 genome; circle 6 (purple), dilution sites in DK1622 genome. The vacuum regions of the top 100 enrichment sites are marked by the outer bold red arrows. The enrichment sites and dilution sites for verification are marked with inner red and purple thin arrows, respectively. (c) Comparison of the local GC content of theoretical TA sites, enrichment sites and dilution sites. *** Wilcoxon test, $p < 0.001$.

3.4. Chromosomal distribution of HEISs in M. xanthus

As shown in Fig. 4b, the 1095 enrichment sites (marked as the inner red circle) were scattered throughout the genome, with slight aggregation in two continuous regions near *oriC*, i.e., the 8.5–9.1 Mb region and the 0–0.7 Mb region, where included 24% of the enrichment sites. The top 100 enrichment sites (highlighted in outer red circle) were also slightly aggregated in the *oriC* region. These results suggested that the gene dosage effect [33] played a role in the enrichment of HEISs, but seemed not to be the main decisive factor. We noticed that there were some vacuum regions for the top enrichment sites, for example, the 0.8–1.4 Mb and 5.5–6.1 Mb regions (marked by outer bold red arrows in Fig. 4b). These two regions cover 418 genes involving in a variety of cellular life processes, and being significantly enriched in metabolic pathways such as secondary metabolite synthesis, biofilm formation and bacterial secretion transport (Figs. S6e and f).

Although the TA contents were significantly different in coding region and intergenic region (Fig. S4a), less than ten percent of the inserted sites (approximately 3.5‰) were identified as enrichment sites in both regions. The similar portion of TA loci also means that the enrichment sites did not bias to distribute in gene coding regions. Analyzing the distribution of these enrichment sites showed that there was no conservative sequence characteristics, but the sites were normally located in the regions with low G + C content. Besides, one of the vacuum regions of enrichment sites (0.8–1.4 Mb) was found to have higher GC content than that of bulk genome (Fig. S7). Comparatively, the dilution sites (the outermost purple circle) were rare in the *oriC* regions (Fig. 4b). Notably, the dilution sites overlapped well with the horizontal transfer genes (green circle), which were also in low GC content (black circle). Thus, although both exhibited significantly lower GC content in local sequences than that of the theoretical TA loci (Wilcoxon test, $p < 0.001$; Fig. 4c), the enrichment sites and the dilution sites exhibited opposite position effects. These dilution sites were not suitable for integration expression of target genes. Similarly, the vacuum regions of enrichment sites should also be avoided for heterologous expression of exogenous genes.

3.5. Experimental determination of M. xanthus HEISs

To verify the expression efficiency of enrichment sites, we selected the top 10 enrichment sites (Table S4) to site-directly integrate the *egfp* gene in the DK1622 genome. The selected sites located in the 3.7–8.8 M region of the chromosome (Fig. 4b), in either the RIR (4 sites) or ORF (6 sites) regions. The 10 associated genes encode a SAM-dependent methyltransferase, a serine/threonine protein kinase, a peptidyl-dipeptidase A, a (2Fe–2S)-binding protein, and hypothetical proteins, respectively. All the above genes were in low transcription according to the transcriptome analysis, and their log (FPKM+1) values were ranged from 0.7 to 2.1, which were below the median (4.8) for the genome-wide transcriptional level (Fig. S2). Insertion of the *egfp* gene into the 10 enrichment sites resulted in the DK-*egfp*-1 to DK-*egfp*-10 mutants. The *egfp* gene was controlled by the moderate Tn5 promoter that was frequently used to express single gene or huge smBGCs in *M. xanthus* [34]. Similarly, six dilution sites were chosen to insert *egfp* by the homologous recombination, resulting mutants DK-*egfp*-11 - DK-*egfp*-16. The DK-19-*egfp* mutant with the *egfp* gene integrated at the *attB* site, the most commonly used site for gene expression in *M. xanthus* [17], was used as control. The insertion sites are marked with inner arrows in Fig. 4c. These mutants all had a similar growth curve (Fig. S8). The mutants with an integration at the enrichment sites all had 15-78 times higher of the *egfp* transcriptional level than that in DK-19-*egfp*, and the fluorescence intensities in DK-*egfp*-1 ~ DK-*egfp*-10 were also 2–5 higher than that in DK-19-*egfp* (Fig. 5a). In comparison, all the six mutants with the insertion of *egfp* at dilution sites exhibited significantly low transcriptional level of *egfp* and fluorescence intensity.

In addition, we also heterologously expressed a *Streptomyces* single-

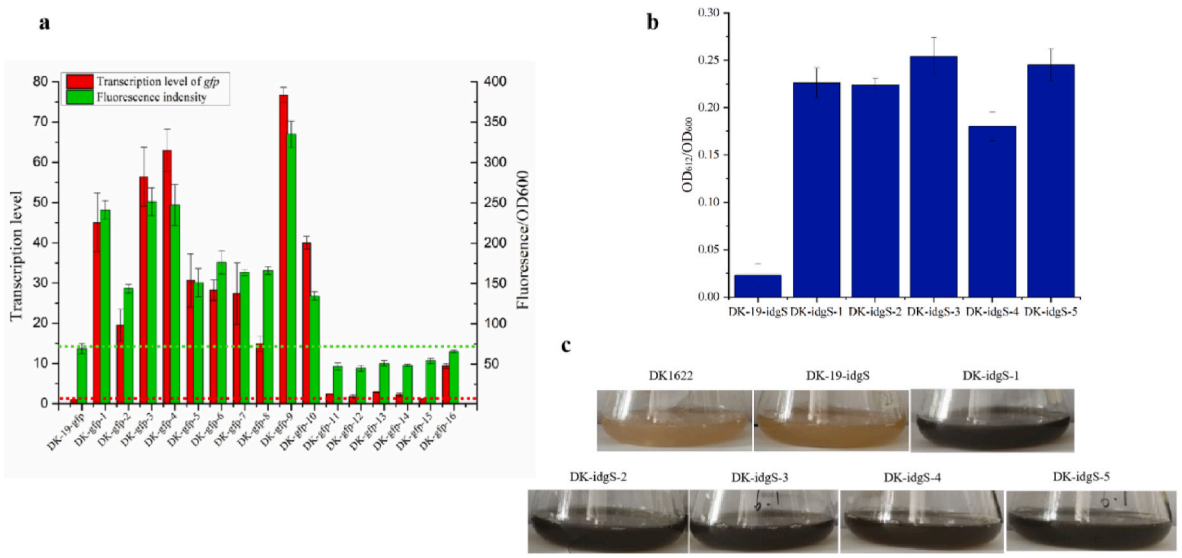

Fig. 5. Verification of the enrichment sites. (a) Analysis of the *egfp* expression at representative enrichment sites and dilution sites. The expression of *egfp* integrated at *attB* site was used as control (the red and green dotted lines represent the transcription and expression of *egfp* in DK-19-*egfp*, respectively). The error bars represent the standard deviation of three independent experiments. (b) Detection of the indigoidine in *M. xanthus* mutants expressing the *idgS*. The error bars represent the standard deviation of three independent experiments. (c) The growth of DK1622 and different *M. xanthus* mutants expressing the *idgS* at representative enrichment sites and *attB* site.

module non-ribosomal peptide synthetase gene *idgS*, which condenses two L-glutamines to blue pigment indigoidine [28,35]. The gene was inserted at either the top 5 enrichment sites or the *attB* site in *M. xanthus* DK1622. Compared with that of the *attB* insertion of *idgS*, the mutants with *idgS* at enrichment sites expressed about 10-fold more indigoidine, making the culture dark or even blue-black (Fig. 5b and **c**). These site-specific expression results demonstrated that the integration sites enriched by high screening pressure represented the candidate loci for the high expression of exogenous genes in *M. xanthus* DK1622.

4. Discussion

Although the chromosomal position effect is known to be commonplace in different hosts, the underlying molecular mechanism remains obscure, and might be with varied reasons. For example, when integrated at different chromosomal loci in *E. coli*, the lactose operon exhibited an increased expression with the proximity to the *oriC*, which was suggested to be the simultaneous DNA replication events leading to a higher copy number of the genes proximal to *oriC* than those nearer to the terminus, i.e., "gene dosage effect" [7,36–39]. However, Bryant et al. inserted the *egfp* at 14 sites in non-coding regions across the *E. coli* chromosome and found varied expression levels of *egfp*, which were not consistent with the gene dosage [40]. These results suggest contingency in experiments when a few or dozens of insertion sites were analyzed, and the aim of this work is to search for plug-to-play HEISs throughout the genome. In this study, we globally analyzed the expression of *egfp* reporter gene integrated along the chromosome of DK1622. The aggregation of enrichment sites and dilution sites in the *oriC* and *Ter* region, respectively, suggests not only the presence of gene dosage effect in *M. xanthus*, but also a tier of regulation on position effect in addition to gene dosage effect. The *Ter* region was reported as a constrained macrodomain structurally distinct from the rest of the chromosome in the circular *E. coli* genome, which may lead to some high-order regulations [41].

Actually, Vora et al. found that the transcriptionally silent extended protein occupancy domains (tsEPODS), which is coincided with the regions that are able to bind with abundant nucleoid-associated proteins (NAPs) [42], was likely to contribute to the transcriptional silencing of inserted *egfp* genes. Some studies also reported that the positions where the expression of *egfp* was unusually high were corresponded to the regions with unusual susceptibility to DNA gyrase inhibitors, suggesting that hyper-expression may be due to increased DNA supercoiling [43,44]. However, if the *gfp* gene was integrated into similar positions of the *E. coli* genome, their expression levels might also be substantially different [7,40,45]. With the 3D structure of *Streptomyces* linear chromosomes revealed, Lioy et al. discovered that most secondary metabolic genes are located at the extremities of the chromosome arms and highly transcribed during the stable stage [32]. We also observed a similar aggregation of SMGCs near the *Ter* in *M. xanthus*, suggesting potential high-order regulation mechanisms for their expressions. Thus, position effect might occur for a variety of reasons and need to be further explored from multiple dimensions.

Nevertheless, the chromosomal position effect does be commonplace in different hosts and can significantly affect the transgene expression, which may provide an efficient strategy to achieve efficient inauguration of target genes in microbial cell factories. Normally, researchers have to screen lots of random insertion mutants to reveal which sites have positive position effects [10,11]. The critical question is how to determine the positive sites for the transgene expression. In this study, we introduced the insertion index to reflect the site insertion abundance, and used enrichment and dilution of insertion index to define the sites with positive and negative position effects, respectively, thus to select the most efficient sites, as well as to avoid false positive. The site-directed insertion of *egfp* or *idgS* demonstrated that the dilution sites, such as horizontal transfer regions, are unsuitable for efficient transgene expression, while the enrichment sites are HEISs, such as the *oriC* region, which can be employed as the plug-and-play sites for efficient expression of integrated genes. Such regional position effects and the related regulation mechanisms need to be further investigated.

5. Conclusion

The production variation from position effect shows applicable for efficient heterologous expression. However, researchers must screen lots of random insertion mutants to obtain the high expression strains. Characterizing HEISs for plug-and-play application is intriguing for efficient expression of alien genes that are integrated in a bacterial chromosome. Thousand-fold expression variation caused by position

effect probably suggests complicated chromosomal spatial organization and effects in *M. xanthus*, which requires further investigations.

Data availability

The sequencing data of RNA-seq and Tn-seq were deposited in NCBI as BioProject PRJNA1010866.

CRediT authorship contribution statement

Xin-jing Yue: Conceptualization, Methodology, Writing – original draft, Funding acquisition. **Jia-rui Wang:** Investigation. **Jun-ning Zhao:** Investigation. **Zhuo Pan:** Software. **Yue-zhong Li:** Conceptualization, Writing – review & editing, Resources, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

We thank Dr. Yihua Chen, State Key Laboratory of Microbial Resources, Institute of Microbiology, Chinese Academy of Sciences, Beijing, China for providing the plasmid pCIMt002. This research was funded by the National Key Research and Development Program of China (2021YFC2101000), and the National Natural Science Foundation of China (32301220).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.synbio.2024.04.007.

References

[1] Feuerborn A, Cook PR. Why the activity of a gene depends on its neighbors. Trends Genet 2015;31:483–90. https://doi.org/10.1016/j.tig.2015.07.001.
[2] Sturtevant AH. The effects of unequal crossing over at the bar locus in Drosophila. Genetics 1925;10:117–47.
[3] Spielmann M, Lupianez DG, Mundlos S. Structural variation in the 3D genome. Nat Rev Genet 2018;19:453–67. https://doi.org/10.1038/s41576-018-0007-0.
[4] Akhtar W, de Jong J, Pindyurin AV, Pagie L, Meuleman W, de Ridder J, Berns A, Wessels LF, van Lohuizen M, van Steensel B. Chromatin position effects assayed by thousands of reporters integrated in parallel. Cell 2013;154:914–27. https://doi.org/10.1016/j.cell.2013.07.018.
[5] Chen X, Zhang J. The genomic landscape of position effects on protein expression level and noise in yeast. Cell Syst 2016;2:347–54. https://doi.org/10.1016/j.cels.2016.03.009.
[6] Wu XL, Li BZ, Zhang WZ, Song K, Qi H, Dai JB, Yuan YJ. Genome-wide landscape of position effects on heterogeneous gene expression in. Biotechnol Biofuels 2017;10. ARTN18910.1186/s13068-017-0872-3.
[7] Block DH, Hussein R, Liang LW, Lim HN. Regulatory consequences of gene translocation in bacteria. Nucleic Acids Res 2012;40:8979–92. https://doi.org/10.1093/nar/gks694.
[8] Cooke K, Browning DF, Lee DJ, Blair JMA, McNeill HE, Huber D, Busby SJW, Bryant JA. Position effects on promoter activity in Escherichia coli and their consequences for antibiotic-resistance determinants. Biochem Soc Trans 2019;47:839–45. https://doi.org/10.1042/Bst20180503.
[9] Jeong DE, So Y, Park SY, Park SH, Choi SK. Random knock-in expression system for high yield production of heterologous protein in Bacillus subtilis. J Biotechnol 2018;266:50–8. https://doi.org/10.1016/j.jbiotec.2017.12.007.
[10] Bilyk B, Horbal L, Luzhetskyy A. Chromosomal position effect influences the heterologous expression of genes and biosynthetic gene clusters in *Streptomyces albus* J1074. Microb Cell Factories 2017;16:5. https://doi.org/10.1186/s12934-016-0619-z.
[11] Zhu LP, Yue XJ, Han K, Li ZF, Zheng LS, Yi XN, Wang HL, Zhang YM, Li YZ. Allopatric integrations selectively change host transcriptomes, leading to varied expression efficiencies of exotic genes in Myxococcus xanthus. Microb Cell Factories 2015;14:105. https://doi.org/10.1186/s12934-015-0294-5.
[12] Peng R, Wang Y, Feng WW, Yue XJ, Chen JH, Hu XZ, Li ZF, Sheng DH, Zhang YM, Li YZ. CRISPR/dCas9-mediated transcriptional improvement of the biosynthetic gene cluster for the epothilone production in Myxococcus xanthus. Microb Cell Factories 2018;17:15. https://doi.org/10.1186/s12934-018-0867-1.
[13] Yue XJ, Cui XW, Zhang Z, Hu WF, Li ZF, Zhang YM, Li YZ. Effects of transcriptional mode on promoter substitution and tandem engineering for the production of epothilones in Myxococcus xanthus. Appl Microbiol Biotechnol 2018;102:5599–610. https://doi.org/10.1007/s00253-018-9023-4.
[14] Yue XJ, Cui XW, Zhang Z, Peng R, Zhang P, Li ZF, Li YZ. A bacterial negative transcription regulator binding on an inverted repeat in the promoter for epothilone biosynthesis. Microb Cell Factories 2017;16:92. https://doi.org/10.1186/s12934-017-0706-9.
[15] Wang Y, Yue XJ, Yuan SF, Hong Y, Hu WF, Li YZ. Internal promoters and their effects on the transcription of operon genes for epothilone production in Myxococcus xanthus. Front Bioeng Biotechnol 2021;9:758561. https://doi.org/10.3389/fbioe.2021.758561.
[16] Huo L, Hug JJ, Fu C, Bian X, Zhang Y, Muller R. Heterologous expression of bacterial natural product biosynthetic pathways. Nat Prod Rep 2019;36:1412–36. https://doi.org/10.1039/c8np00091c.
[17] Yue XJSD, Zhuo L, Li YZ. Genetic manipulation and tools in myxobacteria for the exploitation of secondary metabolism. Engineering Microbiology 2023. https://doi.org/10.1016/j.engmic.2023.100075.
[18] Hodgkin J, Kaiser D. Cell-to-cell stimulation of movement in nonmotile mutants of Myxococcus. Proc Natl Acad Sci U S A 1977;74:2938–42. https://doi.org/10.1073/pnas.74.7.2938.
[19] Yang X, Liu D, Liu F, Wu J, Zou J, Xiao X, Zhao F, Zhu B. HTQC: a fast quality control toolkit for Illumina sequencing data. BMC Bioinf 2013;14:33. https://doi.org/10.1186/1471-2105-14-33.
[20] Li R, Li Y, Kristiansen K, Wang J. SOAP: short oligonucleotide alignment program. Bioinformatics 2008;24:713–4. https://doi.org/10.1093/bioinformatics/btn025.
[21] Dong MJ, Luo H, Gao F. Ori-finder 2022: a comprehensive web server for prediction and analysis of bacterial replication origins. Dev Reprod Biol 2022;20:1207–13. https://doi.org/10.1016/j.gpb.2022.10.002.
[22] Zhao JY, Zhong L, Shen MJ, Xia ZJ, Cheng QX, Sun X, Zhao GP, Li YZ, Qin ZJ. Discovery of the autonomously replicating plasmid pMF1 from Myxococcus fulvus and development of a gene cloning system in Myxococcus xanthus. Appl Environ Microbiol 2008;74:1980–7. https://doi.org/10.1128/AEM.02143-07.
[23] van Opijnen T, Bodi KL, Camilli A. Tn-seq: high-throughput parallel sequencing for fitness and genetic interaction studies in microorganisms. Nat Methods 2009;6:767–72. https://doi.org/10.1038/nmeth.1377.
[24] Brosius J. Plasmid vectors for the selection of promoters. Gene 1984;27:151–60. https://doi.org/10.1016/0378-1119(84)90136-7.
[25] Hu WF, Niu L, Yue XJ, Zhu LL, Hu W, Li YZ, Wu CS. Characterization of constitutive promoters for the elicitation of secondary metabolites in myxobacteria. ACS Synth Biol 2021;10:2904–9. https://doi.org/10.1021/acssynbio.1c00444.
[26] Dornan J, Grey H, Richardson JM. Structural role of the flanking DNA in mariner transposon excision. Nucleic Acids Res 2015;43:2424–32. https://doi.org/10.1093/nar/gkv096.
[27] van Opijnen T, Camilli A. Genome-wide fitness and genetic interactions determined by Tn-seq, a high-throughput massively parallel sequencing method for microorganisms. Curr Protoc Microbiol 2010. https://doi.org/10.1002/9780471729259.mc01e03s19 [Chapter 1]:Unit1E 3.
[28] Li P, Li J, Guo Z, Tang W, Han J, Meng X, Hao T, Zhu Y, Zhang L, Chen Y. An efficient blue-white screening based gene inactivation system for Streptomyces. Appl Microbiol Biotechnol 2015;99:1923–33. https://doi.org/10.1007/s00253-014-6369-0.
[29] Lai FL, Gao F. GC-Profile 2.0: an extended web server for the prediction and visualization of CpG islands. Bioinformatics 2022;38:1738–40. ARTNbtab86410.1093/bioinformatics/btab864.
[30] Kono N, Arakawa K, Tomita M. Comprehensive prediction of chromosome dimer resolution sites in bacterial genomes. BMC Genom 2011;12. Artn1910.1186/1471-2164-12-19.
[31] Bentley SD, Chater KF, Cerdeno-Tarraga AM, Challis GL, Thomson NR, James KD, Harris DE, Quail MA, Kieser H, Harper D, et al. Complete genome sequence of the model actinomycete Streptomyces coelicolor A3(2). Nature 2002;417:141–7. https://doi.org/10.1038/417141a.
[32] Lioy VS, Lorenzi JN, Najah S, Poinsignon T, Leh H, Saulnier C, Aigle B, Lautru S, Thibessard A, Lespinet O, et al. Dynamics of the compartmentalized Streptomyces chromosome during metabolic differentiation. Nat Commun 2021;12:5221. https://doi.org/10.1038/s41467-021-25462-1.
[33] Chandler MG, Pritchard RH. The effect of gene concentration and relative gene dosage on gene output in Escherichia coli. Mol Gen Genet 1975;138:127–41. https://doi.org/10.1007/BF02428117.
[34] Fu J, Wenzel SC, Perlova O, Wang J, Gross F, Tang Z, Yin Y, Stewart AF, Muller R, Zhang Y. Efficient transfer of two large secondary metabolite pathway gene clusters into heterologous hosts by transposition. Nucleic Acids Res 2008;36:e113. https://doi.org/10.1093/nar/gkn499.
[35] Muller M, Auslander S, Auslander D, Kemmer C, Fussenegger M. A novel reporter system for bacterial and mammalian cells based on the non-ribosomal peptide indigoidine. Metab Eng 2012;14:325–35. https://doi.org/10.1016/j.ymben.2012.04.002.
[36] Helmstetter CE. DNA synthesis during the division cycle of rapidly growing Escherichia coli B/r. J Mol Biol 1968;31:507–18. https://doi.org/10.1016/0022-2836(68)90424-5.
[37] Beckwith JR, Signer ER, Epstein W. Transposition of the lac region of E. coli. Cold Spring Harbor Symp Quant Biol 1966;31:393–401. https://doi.org/10.1101/sqb.1966.031.01.051.

[38] Wolf B, Pato ML, Ward CB, Glaser DA. On the origin and direction of replication of the E. coli chromosome. Cold Spring Harbor Symp Quant Biol 1968;33:575–84. https://doi.org/10.1101/sqb.1968.033.01.064.
[39] Sousa C, deLorenzo V, Cebolla A. Modulation of gene expression through chromosomal positioning in Escherichia coli. Microbiology-Sgm 1997;143:2071–8. https://doi.org/10.1099/00221287-143-6-2071.
[40] Bryant JA, Sellars LE, Busby SJ, Lee DJ. Chromosome position effects on gene expression in Escherichia coli K-12. Nucleic Acids Res 2014;42:11383–92. https://doi.org/10.1093/nar/gku828.
[41] Lioy VS, Cournac A, Marbouty M, Duigou S, Mozziconacci J, Espeli O, Boccard F, Koszul R. Multiscale structuring of the E. coli chromosome by nucleoid-associated and condensin proteins. Cell 2018;172:771. https://doi.org/10.1016/j.cell.2017.12.027.
[42] Vora T, Hottes AK, Tavazoie S. Protein occupancy landscape of a bacterial genome. Mol Cell 2009;35:247–53. https://doi.org/10.1016/j.molcel.2009.06.035.
[43] Ma J, Wang MD. DNA supercoiling during transcription. Biophys Rev 2016;8:75–87. https://doi.org/10.1007/s12551-016-0215-9.
[44] Dorman CJ, Dorman MJ. DNA supercoiling is a fundamental regulatory principle in the control of bacterial gene expression. Biophys Rev 2016;8:209–20. https://doi.org/10.1007/s12551-016-0205-y.
[45] Urtecho G, Tripp AD, Insigne KD, Kim H, Kosuri S. Systematic dissection of sequence elements controlling sigma70 promoters using a genomically encoded multiplexed reporter assay in Escherichia coli. Biochemistry 2019;58:1539–51. https://doi.org/10.1021/acs.biochem.7b01069.